Journal of Computer Science 5 (10): 680-689, 2009
ISSN 1549-3636

A Survey of Compute Intensive Algorithms for Ribo Nucleic Acids Structural Detection

Ra'ed Al-Khatib, Rosni Abdullah and Nur'Aini Abdul Rashid
School of Computer Sciences, University Sains Malaysia, 11800 Penang, Malaysia

Abstract: Problem statement: Finding an accurate RNA structural alignment from primary sequence due to it is time consuming and computationally NP-hard problem is a major bioinformatics challenge. According to our investigation majority of current researches were concerned on achieving faster execution time, improving space complexity and better cache management. Recently one research introduced cache-efficient Chip Multiprocessor (CMP) algorithms with good speed-up to exploit parallelism in detection the critical path length. Our contribution in this article was a comprehensive survey of methods for solving RNA secondary structure prediction with Pseudoknots (PK) and sequence alignment in bioinformatics. The aim was to highlight the challenges related issues which would provide sufficient information to assist the new coming researchers in this field as well as a good reference guide for bioinformatics professionals. **Approach:** We computed various algorithms that predicted an RNA molecules secondary structure from primary sequence, without pseudoknots from one side and pseudoknotted RNA secondary structure in the other side. Furthermore, we also reviewed and compared in two tables the methods that developed for RNA structural predictions. **Results:** Our findings of this survey confirmed that Dynamic Programming (DP) method via CMP algorithms can be used to predict the RNA secondary structure with simple PK and it gives good results. **Conclusion:** The methods for predicting RNA's structural are coming in two groups: Firstly, pseudoknotted RNA structural problem is computationally complex and secondly, common methods significantly gave not accurate enough results for predicting pseudoknotted RNA.

Key words: Bioinformatics, RNA secondary structure, pseudoknots, dynamic programming, NP-complete

INTRODUCTION

Bioinformatics is a computer application to manage the biological information and it uses computer to gather, store, analyze, manipulate, interpret and integrate biological, genetic information and macromolecules (Deoxyribo Nucleic Acid (DNA), Ribo Nucleic Acids (RNA), or proteins). One of the most on touched problem is to predict the three-dimensional (3D) RNA structure from the primary sequence. Nowadays, it is still a great challenge for biologist to understand RNA's functionalities, which depend on RNA 3D structural features. The main two experimental methods for structure determination are: The Nuclear Magnetic Resonance (NMR) and the Computational X-ray crystallography, which are a completely accurate method for determining the folded structure of RNA molecule[1]. But unfortunately, both NMR and crystallography are time consuming and very expensive experiments. High level of knowledge is needed to run the experiments which is lacking in the young scientists to overcome this problem[2]. Therefore, three different categories of computational methods to predict the structure of RNA were proposed, (i) Thermodynamic optimal structure or energy minimization model, (ii) comparative sequence and (iii) structure inferring methods[3]. However, these computational methods only provide approximate RNA structural models.

Proteins are an important part of nutrition (diet) to get the proper functioning of the body. Most of the dry weight of the human body and the bodies of other animals is made of protein. RNA molecules an essential ingredient to the synthesis of protein, RNA via messenger RNA (mRNA) type is transcribed from DNA and plays a central role in living cells. According to the central dogma of biology, mRNA is the intermediate carrier of genetic information between DNA and Protein Eq. 1 in a natural process called RNA interference (RNAi) that occurs to regulate the translation of genetic information into proteins. Scientists and researchers have great interest in using

Corresponding Author: Ra'ed Al-Khatib, School of Computer Sciences, University Sains Malaysia, 11800 Penang, Malaysia

this process to create new medications and drugs by using Non-coding RNA (ncRNA) type. Main researches, utilize from RNAi[4], look to discover treatment for: (i) Human Immunodeficiency Virus (HIV) that causes Acquired Immunodeficiency Syndrome (AIDS) and (ii) genital herpes virus or Human Papilloma Virus (HPV) that affects many hundreds million people worldwide by Herpes:

$$\text{DNA} \xrightarrow{\text{Transcription}} \text{RNA} \xrightarrow{\text{Translation}} \text{PROTEIN} \qquad (1)$$

RNA definition and structure: Due to this medical evaluation, mentioned above, for treating or preventing many RNA-related diseases, it's important to know the RNA molecule physical properties and functionality, to understand the function of RNA molecules, we need to understand their structure. RNA molecules definition are: A linear polymer single-stranded chain of alternating phosphate and ribose units Fig. 1a, which its chemical structural consist of a ribose (five-carbon sugar numbered 1' through 5'), a nitrogen-containing base and two phosphate groups, which are attached to the ribose unit, one to the 3' position of ribose and the other to the 5' position. Phosphate groups with ribose sugar units composed the RNA molecules backbone. The nitrogen-containing base may Adenine (A), Guanine (G), Cytosine (C) or Uracil (U) bonded and attached to the 1' position ribose Fig. 1b. Also, each phosphate group has a negative charge at physiological pH (pH: Is a measure of the acidity or basicity of a solution), this negative charge making RNA molecule a charged molecule this mean not stable. The RNA molecule loop structure is a building block for larger structural motifs such as cloverleaf structures Fig. 1c[5].

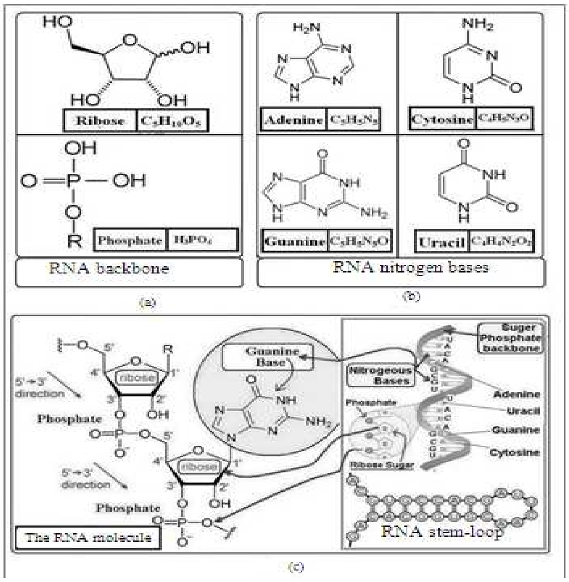

Fig. 1: RNA chemical structure (a) RNA backbone: Phosphate group and ribose (five-carbon sugar) (b) Four nitrogenous bases: Adenine, guanine, cytosine and uracil abbreviated as {A, C, G, U} respectively (c) Chemical structure of RNA molecule

RNA secondary structure (determining the RNA Secondary structure) that can pair up according to the rules in WW:{(A,U),(U,A),(G,C),(C,G),(G,U),(U,G)} Watson-Crick base pairs (G ≡ C) and (A = U) and a Wobble base pair (G-U) to form a triple-, double-, or single-hydrogen bond respectively, which called valid canonical base pairs[6]. So the secondary structure of an RNA molecule is formed by base pairing between various regions of the RNA that result in a configuration of double-helical regions (stems) and single stranded loops, thus it is the collection of base pairs. Given an RNA sequence with primary structure = {A-G-G-C-C-U-U-C-C-U}, using the WW-folding to understand the RNA secondary structure, we can expect six stem loops. Figure 2 explains these six stem loops.

The thermodynamic hypothesis of the actual secondary structure of RNA sequence is the one with the Minimum Free Energy (MFE) such as the base-pairs will increase the structural stability. But unpaired bases decrease that. Our goal, is to calculate the free energy of RNA secondary structure by calculate the total of the energies of all base pairs by taking account that the energy for G ≡ C, A = U and G-U are different this is summarized in equation2 [7]; it minimizes the total free energy:

$$\text{Total MFE for RNA} = \sum \text{of Loop Energies} \qquad (2)$$
$$\text{(At fixed temperature + ionic concentration)}$$

Hence, the task and function of the RNA cannot be determine by secondary structure prediction alone as shown in Fig. 3; the prediction accuracy of RNA structure with the MFE method alone is usually not high, because the energy model is not accurate enough and RNA may not fold into MFE always.

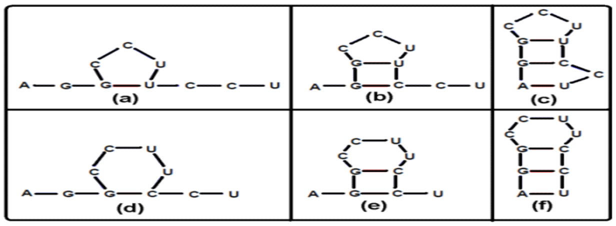

Fig. 2: Six-expectation possible for sequence RNA = {A-G-G-C-C-U-U-C-C-U} by using WW-folding rules

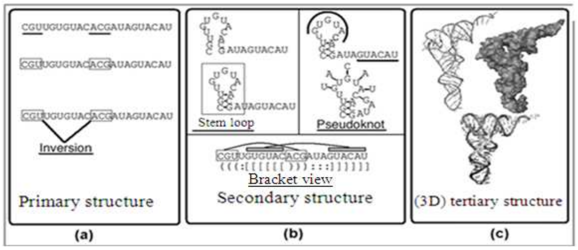

Fig. 3: RNA Structures: (a) RNA Primary Structure (b) RNA secondary structure: (Stem-loop left and pseudoknot right) (c) RNA Three-Dimensional (3D)/tertiary structure

Also, the secondary structure of an RNA sequence must contain multiple loops to be stable in MFE. These single-stranded stem loops can be divided into two large groups: Stem-loops and pseudoknots as shown in Fig. 3b[8]. RNA Pseudoknot structure exist if the RNA 3D secondary structure contains two stem-loop crossing stems or more, in fact, pseudoknots are found in almost all classes of RNA, especially in the genomes of some viruses, as a result we have to use a suitable widespread motif algorithms (strategy) for RNA structural prediction problems, this strategy should take into account the 3D RNA Secondary structure prediction with Pseudoknots and MFE (stable) and often closely related with the biological functions of an RNA sequence[4].

The tertiary structure (3D) is the complete form for RNA folded molecules enabling them to perform their functional role in the cell and is often the key to its function Fig. 3c. Generally, three-dimensional form of RNA sequences is called: 3D functional structure which characteristics are important in biology; firstly, RNA 3D structures are critical to their biological functions, secondly, RNA 3D structures properties may also help identify subsequences of nucleotides that interact with other molecules or complexes.

Consequently, in last decade, predicting the structure of RNA secondary structure prediction with simple pseudo-knots based on minimum free energy (RNA-SP based on MFE) has become biological and medical demands because RNA molecule has two important functions: Regulatory processes to the synthesis of proteins and viral replication, which it is found important in antiviral treatment design[4].

Roadmap: After highlighting the fundamental RNA definition, chemical structure and RNA (Primary, Secondary and tertiary 3D) structures, the basic concept for RNA secondary structure problem.

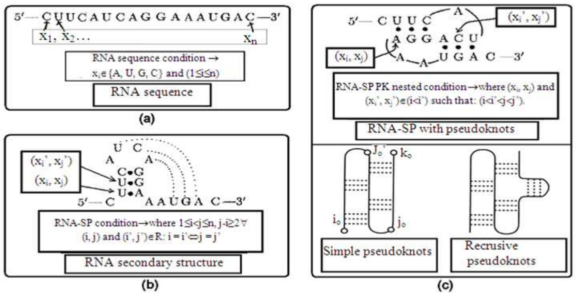

Fig. 4: RNA definition (a) RNA sequence (b): RNA secondary structure (c) RNA-SP with PseudoKnot and down two types of PK (simple and recursive)

Then we classify RNA methods into two groups; at first, methods that consider RNA stem-loops (w/o pseudoknots), secondly, methods for prediction RNA secondary structure with pseudoknots. Next we compare the results for the main methods. Finally, we give some concluding remarks and we present our future plan.

Problem domain: Predicting and producing RNA secondary structure from the sequence is important to understand RNA functions Eq. 3. The RNA fold recognition methods attempt to predict the accurate and more stable RNA folding structure with MFE. RNA 3D structure, in some parts, takes pseudoknots folding Fig. 4.

$$\text{RNA Sequence} \rightarrow \text{RNA-SP Structure} \rightarrow \text{RNA Function} \quad (3)$$

We will define RNA-SP with pseudo-knot as follows:

- RNA sequence is viewed as a string of n characters $x_i = x_1x_2\ldots x_n$ where $xi\in\{A,U,G,C\}$ the four bases and $1\leq i\leq n$ as shown in Fig. 4a
- A single-stranded RNA secondary structure is a list of base-pairs can be viewed as a set[9], X, form an admissible base pairs (x_i, x_j) where at first, $1\leq i<j\leq n$, secondly, $j-i>t$ where t is a small constant, i.e. $j- i \geq 2$. For all base pairs (x_i, x_j) and (x_i', x_j') in X, i = i' if and only if j = j', (i.e., such that $\forall$ (i, j), $(i',j')\in R$: $i = i' \Leftrightarrow j = j'$) as shown in Fig. 4b, this means; two bases that form a pair must be located at different locations, the sequence doesn't fold too sharply on itself and each base can be paired with at most one base, respectively. Also, we allowed just WW base pairs:{(A,U), (C,G), (G,U)}
- RNA include pseudoknot in X is viewed if and only if there exist base pairs (x_i, x_j), $(x_i', x_j')\in X$

(i<i') such that i<i'<j<j' (nested condition) Fig. 4c up. We can find types of pseudoknot: (simple or recursive) Fig. 4c down[9]. So, a given RNA sequence X can with maximum number of base pairs and exponential number of possible structures, Addition to the compute an RNA structure with Minimum Overall Free Energy (MFE)

These complicated motifs contribute to make the general RNA secondary structure with pseudoknots prediction problem are an NP-Complete Problem, because the algorithms for solving an RNA-SP with PK prediction problem need to allow energy functions and it runs in worst case polynomial time. In fact, [9,10] proved that finding pseudoknotted RNA structure with MFE is NP-hard problem, particularly by applying the standard nearest-neighbor energy function. So, researchers of pseudoknotted RNAs are facing with three problems: First, RNA secondary structure prediction with pseudoknots is high cost computationally in run-time and memory space, which made the problem to be NP-complete problem[11] and most professional algorithms exist only for partial classes of pseudoknots, not for all kinds. Second, almost all main RNAs computational methods have been analyzed nested RNA-SP structure, either neglecting RNA pseudoknots for simplicity, or they did not know the pseudoknots side[12]. And lastly, existing RNA prediction programs are suffered from low quality and they are not very reliable.

MATERIALS AND METHODS

Overview of RNA secondary structure algorithms and methods: predicting RNA secondary structure nowadays becomes very important task in bioinformatics. Various works and many researchers made many efforts or introduced several techniques, methods and algorithms for solving RNA-SP problem, these researches can be divided into two main parts as follows:

Solving RNA stem-loops group: This group of research did not consider pseudoknots in solving RNA-SP problem. For more simplicity, they neglected pseudoknots in their study for predicting RNA structure. Many methods and techniques have been implemented for solving RNA secondary structure predictions in the last three decades. Reducing run-time and space complexities and guarantying to give the MFE structure based on the free energy evaluation and thermodynamic models, but not always the lowest MFE is the correct structural RNA molecules fold. In 1978, Waterman and Smith[13,14] and Nussinov *et al.*[15] proposed a first simplified thermodynamic energy model using Dynamic Programming (DP) algorithms to predict RNA secondary structure. They presented DP algorithms which required $O(n^3)$ run-time steps and $O(n^2)$ space complexity, where n length of an RNA sequence.

Many researchers attempt to improve the DP based algorithms used in RNA secondary structure prediction[16-21]. Among these DP algorithms Zuker's Algorithm[16] is the most popular one, this algorithm explored all possible unpseudoknotted RNA secondary structure based on thermodynamic energy minimization model and required $O(n^3)$ run-time and $O(n^2)$ space complexities, where n is the length of an input RNA sequence. MFOLD[22] and ViennaRNA[23] packages implemented with Zuker's DP algorithm. Another approach for large RNAs was introduced by Eddy[17] used divide and conquer strategy. Eddy utilized Myers/Miller algorithm[24], Eddy algorithm was a DP solution runs in $O(n^2 \log n)$ space complexity and made an optimal structural alignment of large RNAs with reducing the memory requirement of Stochastic Context Free Grammar (SCFG) alignments. A main Parallel DP algorithm for detecting pseudoknot-free secondary structure of an RNA molecules was introduced by Tan *et al.*[18], which implemented on NUMA cluster systems by using sequential DP Algorithm and it needs $O(\frac{n^4}{P})$ run-time and $O(\frac{n^3}{P})$ space in cluster, where P is the number of processors and n is a length of RNA sequence.

Dynamic programming approaches for RNA prediction suffer from high computational running time and computing an optimal solution based on MFE in thermodynamic model. Due to these reasons many heuristic methods were proposed. STRAL was recently presented as a heuristic method for alignment of ncRNAs by Dalli *et al.*[19], which is a multiple RNA alignment program that combines structural and sequence information in a 'cheap' DP Algorithms and a heuristic method for mainly alignment of ncRNAs. STRAL needs $O(k^2n^2)$ run-time and $O(n^2)$ memory cost, where n is a length of RNA sequence and k is the matching bases from different two sequences, because STRAL is a heuristic method that reduces sequence structure alignment to a two-dimensional (2D) problem similar to standard multiple sequence alignment. Ideally, an ncRNAs are RNA molecules that do not code for proteins, but ncRNA are important for functional in biological processes, including localization, replication, translation, degradation and stabilization of biological macromolecules. Next, the previous Sparse Dynamic Programming (SDP) approach was used and improved from Ogurtsov *et al.*[20]. This was finding the optimal Multi-Branch Loop-Free (MLF) structure for evaluating and internal loops. SDP algorithm implemented in Afold tool and it has run-time of $O(M*\log^2 L)$ and work space of O(M), where $M<L^2$ is the number of possible nucleotide pairings and L is the length of an RNA molecule. It was improved on Lyngsø *et al.*[25] earlier study which time was reduced from $O(n^4)$-$O(L^3)$ or $O(n^3)$, who used DP algorithms to find the RNA-SP with MFE and analysis internal loops.

Recently, a Co-folding DP Algorithm was developed by Ziv-Ukelson *et al.*[21], that obtained run-time $O(n^4\zeta(n))$, where $\zeta(n)$ can converge to O(n), markedly it was developed from Sankoff's dynamic programming algorithm from[26], Sankoff's algorithm requires $O(n^6)$ time and $O(n^4)$ space. And up to date, Mathuriya *et al.*[6] presented GTfold which is a parallel implementation multicore and scalable program for RNA-SP without Pseudoknots.

Solving RNA with pseudoknots group: All the algorithms discussed in this part consider pseudoknots in their works. In introduction, we gave a convinced reason that folding pseudoknots in RNA-SP perform essential functions in both: (i) as part from transcription machinery in cell for proteins synthesis and regulatory processes. (ii) as part from antiviral drug design because RNA activities have important results here[27]. Many researchers and study gave various techniques in RNA-SP with Pseudoknots; such as Pleij *et al.*[28] the first general method for Plausible RNA folding with pseudoknots, while RNA with pseudoknots noted and coined before[16,29]. Abrahams *et al.*[30] developed and promoted a local search method by using computer simulation. Van Batenburg *et al.*[31] and Gultyaev *et al.*[32] investigated Genetic Algorithms (GA), while Shapiro and Wu developed a parallel (GA) for detecting H-pseudoknots[33], Lyngsø and Pedersen[10] explained that RNA-SP with pseudoknot structure prediction problem is based on difficult mathematic problems, such as NP-problem and it needs exponential time algorithms. Several earlier study introduced Dynamic Programming (DP) algorithms to find MFE structure for RNA secondary structure prediction with pseudoknots, we index them as follows:

- First DP algorithm to give an optimal lowest energy prediction for RNA structure with pseudoknots called pknotsRE was introduced by Rivas *et al.*[34], which is a complete model for calculating the free energy of pseudoknotted RNA secondary structure. However, pknotsRE demanded high run-time and space complexity of $O(n^6)$ and $O(n^4)$, respectively for RNA sequence of length n, making this algorithm infeasible to run on large RNA molecules. A pknotsRE algorithm has advantages; it considered the first one for determining the MFE and handled large two classes of RNA with pseudoknots; the arbitrary planar class and the restricted non-planar pseudoknots class
- Another method considered the non-recursive class in RNA with simple pseudoknots was presented by Lyngsø and Pedersen[11] using a polynomial-time and space DP algorithm with $O(n^5)$ and $O(n^3)$ of time and space complexity, respectively. They then proved that predicting pseudoknotted RNA secondary structure in general is NP-hard problem. Also, in the same time a polynomial-time and space DP algorithm to compute RNA secondary structures with maximum number of base pairs with presence simple pseudoknots was designed by Akutsu[9], which runs in $O(n^4)$ time and $O(n^3)$ space complexity and $O(\frac{n^4}{B})$ cache-misses, namely cache-misses is the better cache management memory access is determined by if the accessed data block is a cache hit or a cache miss, where B is the memory block size and n is an RNA sequence length
- One partition DP function algorithm called NUPACK for Nucleic Acid was transformed by Dirks and Pierce[35]. NUPACK was extended to include the most physically relevant pseudoknots for the standard secondary structure energy model, it is computing and calculating the partition function of base-pairing probabilities RNA with or w/o pseudoknots and single-stranded DNA (ssDNA) molecules and required $O(n^5)$ run-time and $O(n^4)$ space complexity
- Many reasons leaded Pseudoknotted RNA secondary structure researchers to adopt Heuristic Approaches. These reasons that guided to go to the heuristic methods are; (i) that most of the DP methods are impractical because theirs computational high cost, they required for run-time (from $O(n^4)$ to $O(n^6)$) and for time space complexity (from $O(n^2)$ to $O(n^4)$). (ii) the practical solution needs side. These reasons guide the researchers to go to heuristic part for reducing theirs run-time and space complexities. While many heuristic approaches for predicting pseudoknotted RNA are simulate a hypothetical process of folding, the main early heuristic algorithms are presented[30-32]. The most popular heuristic DP algorithm one called Iterated Loop Matching (ILM) algorithm was

produced by Ruan *et al.*[36]. It was based on stem zone developed for the Loop Matching (LM) algorithm (Nussinov *et al.*[15]). ILM method can predict pseudoknotted RNA for both aligned and individual sequences and can use either thermodynamic or comparative models or both with $O(n^4)$ time and $O(n^2)$ space complexity. ILM is also minimizing free energy model in the average run-time of $O(n^3)$ without changed in space complexity. Subsequently, HotKnots Heuristic algorithm was presented by Ren *et al.*[37], which was out-performed the heuristic ILM algorithm. Recently, pseudoknotted RNA detection Heuristic algorithm called KnotSeeker was presented by Sperschneider and Datta[27], which was used a hybrid sequence matching and Minimum Free Energy (MFE) to obtain more accurate in RNA secondary structure with pseudoknots detection, especially for long sequences. Latest heuristic pseudoknotted RNA detection algorithm was presented by Li[38] to predict main arbitrary RNA including pseudoknots and maximize stems. It required $O(n^3)$ time and $O(n)$ space complexity and it got more improvement results in sensitivity and specificity

- DP algorithm to predict RNA with simple pseudoknots based on using standard thermodynamic parameters was made by Deogun *et al.*[39], it made improvement on Akutsu research[9] in worst case time and space complexities of $O(n^4)$ and $O(n^3)$, respectively
- Extending from[34] pknotsRE Rivas work a good DP algorithm called pknotsRG-mfe was developed by Reeder and Giegerich[40]. A pknotsRG-mfe is an augmented version from pknotsRE and predicting restrict class of simple nested pseudoknotted RNA structure and provided suboptimal structures and it has reduced the run-time and space complexities to $O(n^4)$ and $O(n^2)$, respectively
- A DP algorithm was developed by Li and Zhu[41] developed for predicting RNA including: (nested and subclass of crossed Pseudoknots) with $O(n^4)$ time, $O(n^3)$ space. This algorithm has same power of Rivas Algorithm[34] for predicting the planar pseudoknots and can predict more complex Pseudoknotted RNA comparing with PknotsRG Reeder Algorithm[40], too
- Pseudoknot Local Motif Model and Dynamic Partner Sequence Stacking (PLMM_DPSS) algorithm was introduced by Huang and Ali[42]. PLMM_DPSS algorithm used a modification of Needleman and Wunsch work in the DP for RNA sequence alignment algorithm[43]
- An applicable DP and parallel algorithm for string problem called Cache-Oblivious (CO) algorithm was presented by Chowdhury *et al.*[44], which matched good run-time $O(n^4)$, made improvement in space complexities to $O(n^2)$, gained better cache-missed $O(\frac{n^4}{B\sqrt{M}})$ and the CO algorithm implemented in $O\left(\frac{n^4}{p}+n^2\log n\right)$ parallel steps when executed in P processors, where n is an RNA sequence length, M is a cache of size and B is the memory block size. Also, Chowdhury *et al.*[45] presented new version of CO DP algorithm for solving RNA-SP with pseudoknots prediction, which it made improvement for Akutsu algorithms[9] in space and cache to $O(n^2)$ and $O(\frac{n^4}{B\sqrt{M}})$ respectively, with keeping its time complexity same in $O(n^4)$, where M is a cache of size, B is the memory block size, we know always n is the length of an RNA sequence
- An improvement DP algorithm called Hierarchical Fold (HFold) worked by Jabbari *et al.*[46], it required $O(n^3)$ running time and $O(n^2)$ space complexity. This approach can predict a wide range of biological MFE pseudoknotted RNA secondary structures and made a good improvement in running time;(from $O(n^6)$ to $O(n^3)$), for predicting MFE nested kissing hairpins from the previous well known Algorithm Rivas and Eddy[34]
- A cache-efficient DP Chip Multiprocessor (CMP) algorithm was presented by Chowdhury and Ramachandran[47], this algorithm obtained a good amount of parallelism on cache-efficient critical path. They used and combined this algorithm to serve RNA secondary structure prediction with simple pseudo-knots; they got $O(n^4)$ in sequential running time, $O(\frac{n^4}{B\sqrt{M}})$ in cache-efficiency and $O(n)$ in amount of parallel, this mean they improved in critical path length from their previous study that mentioned in number(9)[44]. Where the variables n is the length of RNA sequence, B is the memory block size

RESULTS AND DISCUSSION

Many researchers made RNA secondary structure predictions w/o pseudoknots methods to solve RNA-SP problem and their consequences were promising as demonstrated in Table 1, for the RNA secondary structure predictions with pseudoknots problem many scientists attempted to solve RNA-SP problem and also they obtained promising results as illustrated in Table 2.

Table 1: RNA secondary structure predictions " Stem-Loops" w/o pseudoknots methods and techniques

No.	Method	Reference and year	Time complexity	Space complexity	Main contribution
1	Co-folding DP alg.	A faster algorithm for RNA Co-folding, Ziv-Ukelson *et al*., 2008[21].	$O(n^4\zeta(n))$	-	A Co-folding DP alg. as faster and based on Sankoff's Alignment SA.
2	SDP Alg. analysis of internal loops in the RNA-SP.	Analysis of internal loops within the RNA secondary structure in almost quadratic time, Ogurtsov *et al*., 2006[20].	$O (M * \log^2 L)$	$O (M)$	A Sparse DP Alg. For optimal MFL structure to analyze the Internal loops in the RNA-SP with NNM energy functions.
3	STRAL: Multiple RNA alignment prog. in a 'cheap' DP Alg.	STRAL: Progressive alignment of non-coding RNA using base pairing probability vectors in quadratic time, Dalli *et al*., 2006[19].	$O (k^2n^2)$	$O(n^2)$	A STRAL: Multiple RNA alignment prog. that combines structural and sequence information in a 'cheap' DP Alg.
4	Parallel DP alg. For RNA-SP.	Load Balancing Algorithm in Cluster-based RNA secondary structure Prediction, Tan *et al*. 2005[18].	$O (n^4/P)$	$O(n^3/P)$	A parallel alg. for RNA-SP in NUMA cluster systems by using sequential DP Algorithm.
5	DP Solution for large RNA-SP by using divide and conquer Alg.	A memory-efficient dynamic programming algorithm for optimal alignment of a sequence to an RNA secondary structure, Eddy 2002[17].	-	$O(n^2 \log n)$	A DP Solution to the RNA-SP problem for a large by using divide and conquer strategy.
6	A method to evaluate internal loops by using energy rules.	Fast evaluation of internal loops in, RNA secondary structure prediction Lyngsø *et al*., 1999[25].	$O (n^3)$	-	A method to find part of structure prediction from RNA by using energy rules to evaluate internal loops.
7	Zuker DP Alg. for RNA sequence with minimum energy structure.	Optimal computer folding of large RNA sequences using thermo -dynamics and auxiliary information, Zuker *et al*., 1981[16].	$O (n^3)$	$O(n^2)$	A DP Alg. for folding non -pseudoknotted RNA sequence with minimum energy structure in thermodynamic model.

Table 2: RNA secondary structure predictions with pseudoknots methods and techniques

No.	Method	Reference and year	Time complexity	Space complexity	Cache-efficiency (I/O complexity)	No. of parallel Step (I ∞)	Main contribution
1	Heuristic Algorithm to predict RNA PK to Max. stems.	Heuristic Algorithm for pseudoknotted RNA structure prediction, Li 2008[38].	$O (n^3)$	$O (n)$	-	-	A heuristic algorithm to predict pseudoknotted RNA structure to max. stems and considering only stacking energy.
2	Cache-oblivious DP Alg.	Cache-oblivious dynamic Programming for bioinformatics, Chowdhury *et al*. 2008[45].	$O (n^4)$	$O (n^2)$	$O\left(\frac{n^4}{\sqrt[B]{M}}\right)$	-	CO Alg. by using DP Alg. for sequence alignment and for RNA-SP with simple PK.
3	Cache-efficient CMP alg. by using DP Alg.	Cache-efficient dynamic programming Algorithms for Multicores, Chowdhury *et al*. 2008[47].	$O (n^4)$	-	$O\left(\frac{n^4}{\sqrt[B]{M}}\right)$	$O (n)$	A cache-efficient CMP Alg. by using DP Alg. it using the seq. RNA-SP Alg. with combining between 3D LDDP and GEP.
4	HFold DP alg. that solved the H-MFE RNA-SP problem.	HFold: RNA pseudoknotted secondary structure prediction using hierarchical folding, Jabbari *et al*. 2007[46].	$O (n^3)$	$O (n^2)$	-	-	HFold DP alg. to solve the H-MFE RNA-SP problem, for the class of density-2.
5	CO- cache-efficient Alg. and parallel	Efficient cache-oblivious string Alg. algorithms for bioinformatics, Chowdhury *et al*. 2007[44].	O(n4)	O(n2)	$O\left(\frac{n^4}{\sqrt[B]{M}}\right)$	$O\left(\frac{n^4}{p}+n^2\log n\right)$	CO framework for DP problems, and applied it to obtain efficient CO Alg.s for RNA-SP with simple PK.
6	DP alg. to predict the optimal RNA -SP including Pk.	A new pseudoknots folding algorithm for RNA structure Prediction,Li and Zhu 2005[41].	$O (n^4)$	$O (n^3)$	-	-	A new DP alg. to predict the RNA-SP including nested and a subclass of crossed PK.
7	PknotsRG-mfe DP alg. for folding RNA -SP including PK under the MFE model.	Design, implementation and evaluation of a practical pseudo-knot Folding algorithm based on thermodynamics, Reeder and Giegerich 2004[40].	$O (n^4)$	$O (n^2)$	-	-	A PknotsRG-mfe DP alg. to predict RNA-SP and consider class of simple recursive PK. it is an augmented version of PknotsRE
8	DP Alg. for RNA -SPwith simple pseudoknots.	RNA secondary structure prediction with simple pseudo -knots, Deogun *et al*. 2004[39].	$O (n^4)$	$O (n^3)$	-	-	A DP Alg. for optimal RNA-SP with simple pseudoknots using standard thermodynamic parameters for RNA folding.
9.	ILM Alg. from on stem zone to predict RNA-SP with PK.	An Iterated loop matching approach heuristic alg. Based to the prediction of RNA secondary structures with pseudoknots, Ruan *et al*. 2004[36].	O(n4) Avg. Case ~O(n3)	O(n2)			A heuristic alg. called: ILM Alg. for reliably and efficiently predicting RNA –SP with PK for both aligned and individual sequences.
10	NUPACK DP Alg. transforms Partition Function of an RNA/ssDNA.	A partition function algorithm for nucleic acid secondary structure including pseudoknots, Dirks and Pierce 2003[35].	$O (n^5)$	$O (n^4)$	-	- .	A nonredundant NUPACK DP partition function Alg. that computes a series of recursion for RNA / ssDNA

Table 2: Continued

11	Alg. for RNA-SP Problem including pseudoknots.	RNA pseudoknot prediction in energy based-models, Lyngsø *et al*. 2001[10].	$O(n^5)$	$O(n^3)$	-	-	Alg. for RNA-SP Problem considered the class of one planar non-recursive PK.
12	A simple DP Alg. For RNA-SP.	Dynamic programming algorithm -ms for RNA secondary structure prediction with pseudoknots, Akutsu 2000[9].	$O(n^4)$	$O(n^3)$	-	-	A DP Alg. For RNA-SP with simple PK. for the number of base pair maximization.
13	Polynomial-time and space DP algorithm. with simple PKs.	Pseudoknots in RNA secondary structures. Lyngsø and Pedersen 2000[11].	$O(n^5)$	$O(n^4)$	$O\left(\frac{n^4}{\sqrt[B]{M}}\right)$	-	A polynomial-time and space DP algorithm to consider the non -recursive class from RNA
14	PknotsRE The First DP Alg. to predict an optimal RNA-PK.	A dynamic programming algorithm for RNA structure prediction including PK's, Rivas and Eddy 1999[34].	$O(n^6)$	$O(n^4)$	-	-	A PknotsRE DP Alg. to predict RNA that can handle a large class of arbitrary planar and restricted non-planar of special PK.

CONCLUSION

In preceding years, several challenges of bioinformatics appeared, main one is the predicting of the correct and accurate RNA secondary structure prediction with pseudoknots from primary sequence alignment. Many methods have been successfully done to solve this problem from computational side. In this study, we present the main general methods can be used for solving RNA-SP problem.

The aim research of this study primarily focuses on two features of RNA structural alignment issue: first are the methods deals with RNA folding and second is that methods solve RNA secondary structure prediction with pseudoknots problem. Hence, the RNA secondary structure problem with simple pseudoknots can be solved by using DP algorithms utilizing parallel computing platform on CMP. Thus, developing an efficient parallelization of DP algorithms with accurate method for predicting RNA secondary structure with pseudoknot will be the prominent idea for our future research.

ACKNOWLEDGEMENT

This research has been funded by University Sains Malaysia short-term Research Grant for "KNOWLEDGE PROCESSING PLATFORM" [1001/PKOMP/817002] and by University Sains Malaysia IPS fellowship.

REFERENCES

1. Lukavsky, P.J., 2007. Basic Principles of RNA NMR Spectroscopy. NATO Secur. Through Sci. Ser., 2006: 65-80. DOI: 10.1007/978-1-4020-5900-1.
2. Cheong, H.K., E. Hwang, C. Lee, B.S. Choi and C. Cheong, 2004. Rapid preparation of RNA samples for NMR spectroscopy and X-ray crystallography. Nucleic Acids Res., 32: 10-84. DOI: 10.1093/nar/gnh081
3. Jansson, J., S. Ng, W. Sung and H. Willy, 2004. A Faster and More Space-efficient algorithm for inferring arc-annotations of rna sequences through alignment. Algorithmica, 46: 323-245. http://portal.acm.org/citation.cfm?id=1165837
4. Brierley, I., S. Pennell and R.J.C. Gilbert, 2007. Viral RNA pseudoknots: Versatile motifs in gene expression and replication. Nat. Rev. Microbiol., 5: 598-610. DOI: 10.1038/nrmicro1704
5. Lee, J.C. and R.R. Gutell, 2004. Diversity of base-pair conformations and their occurrence in rRNA structure and RNA structural motifs. J. Mol. Biol., 344: 1225-1249. DOI:10.1016/j.jmb.2004.09.072
6. Mathuriya, A., D. Bader, C.E. Heitsch and S.C. Harvey, 2009. GTfold: A scalable multicore Code for RNA secondary structure prediction. Proceedings of the 2009 ACM Symposium on Applied Computing, (SAC'09), ACM Press, Honolulu, Hawaii, pp: 981-988. http://doi.acm.org/10.1145/1529282.1529497
7. Mathews, D.H., 2005. Predicting a set of minimal free energy RNA secondary structures common to two sequences. Bioinformatics, 21: 2246-2253. DOI: 10.1093/bioinformatics/bti349
8. Taufer, M., A. Licon, R. Araiza, D. Mireles, F.H.D. Van Batenburg, A.P. Gultyaev and M.Y. Leung, 2009. PseudoBase++: An extension of PseudoBase for easy searching, formatting and visualization of pseudoknots. Nucleic Acids Res., 37: D127-D135. DOI: 10.1093/nar/gkn806
9. Akutsu, T., 2000. Dynamic programming algorithms for RNA secondary structure prediction with pseudoknots. Discrete Applied Math. 104: 45-62. DOI: 10.1016/s0166-218x(00)00186-4
10. Lyngsø, B. and C.N. Pedersen, 2000. RNA pseudoknot prediction in energy-based models. J. Comput. Biol., 7: 409-427. DOI: 10.1089/106652700750050862

11. Lyngsø, R.B. and C.N. Pedersen, 2000. Pseudoknots in RNA secondary structures. Proceeding of the 4th Annual International Conferences on Compututational Molecular Biology (RECOMB00), ACM Press, BRICS Report Series RS-00-1. http://doi.acm.org/10.1145/332306.332551
12. Mohl, M., S. Will and R, Backofen, 2009. Lifting Prediction to Alignment of RNA Pseudoknots. Lecture Notes Comput. Sci., 5541: 285-301 DOI: 10.1007/978-3-642-02008-7_22
13. Waterman, M.S., 1978. Secondary structure of single stranded nucleic acids. Adv. Math. Suppl. Stud., 1: 167-212. http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.15.4425
14. Waterman, M.S. and T.F. Smith, 1978. RNA secondary structure: A complete mathematical analysis. Math. Biosci., 42: 257-266. http://eprints.kfupm.edu.sa/62752/
15. Nussinov, R., G. Pieczenik, J.R. Griggs and D.J. Kleitman, 1978. Algorithms for loop matchings. SIAM. J. Applied Math., 35: 68-82. http://link.aip.org/link/?SMJMAP/35/68/1
16. Zuker, M. and P. Stiegler, 1981. Optimal computer folding of large RNA sequences using thermodynamics and auxiliary information. Nucleic Acids Res., 9: 133-148. http://www.ncbi.nlm.nih.gov/pubmed/6163133
17. Eddy, S.R., 2002. A memory-efficient dynamic programming algorithm for optimal alignment of a sequence to an RNA secondary structure. BMC. Bioinform., 3: 18-18. DOI: 10.1186/1471-2105-3-18
18. Tan, G., S. Feng and N. Sun, 2005. load balancing algorithm in cluster-based RNA secondary structure prediction. Proceeding of the 4th International Symposium on Parallel and Distributed Computing, July 4-6, IEEE Xplore Press, pp: 91-96. DOI: 10.1109/ISPDC.2005.32
19. Dalli, D., A. Wilm, I. Mainz and G. Steger, 2006. STRAL: Progressive alignment of non-coding RNA using base pairing probability vectors in quadratic time. Bioinformatics, 22: 1593-1599. DOI: 10.1093/bioinformatics/btl142
20. Ogurtsov, A.Y., S.A. Shabalina, A.S. Kondrashov and M.A. Roytberg, 2006. Analysis of internal loops within the RNA secondary structure in almost quadratic time. Bioinformatics, 22: 1317-1324. DOI: 10.1093/bioinformatics/btl083
21. Ziv-Ukelson, M., I. Gat-Viks, Y. Wexler and R. Shamir, 2005. A Faster Algorithm for RNA Co-folding. Lecture Notes Comput. Sci., 5251: 174-185. DOI: 10.1007/978-3-540-87361-7_15
22. Zuker, M., 2003. Mfold web server for nucleic acid folding and hybridization prediction. Nucleic Acids Res., 31: 3406-3415. DOI: 10.1093/nar/gkg595
23. Hofacker, I.L., P.F. Stadler, L.S. Bonhoeffer, M. Tacker and P. Schuster, 1994. Fast folding and comparison of RNA secondary structures. Monatshefte für Chemie/Chem. Monthly, 125: 167-188. DOI: 10.1007/BF00818163
24. Myers, E.W. and W. Miller, 1988. Optimal alignments in linear space. Comput. Applied Biosci., 4: 11-17. http://www.ncbi.nlm.nih.gov/pubmed/3382986
25. Lyngsø, R.B., M. Zuker and C.N.S. Pedersen, 1999. Fast evaluation of internal loops in RNA secondary structure prediction. Bioinformatics, 15: 440-445. http://www.ncbi.nlm.nih.gov/pubmed/10383469
26. Sankoff, D., 1985. Simultaneous solution of the RNA folding, alignment and protosequence problems. SIAM J. Applied Mathe. 45: 810-825.
27. Sperschneider, J. and A. Datta, 2008. KnotSeeker: Heuristic pseudoknot detection in long RNA sequences. RNA., 14: 630-640. DOI: 10.1261/rna.968808
28. Pleij, C.W., K. Rietveld and L. Bosch, 1985. A new principle of RNA folding based on pseudoknotting. Nucleic Acid Res., 13: 1717-1731. http://www.ncbi.nlm.nih.gov/pubmed/4000943
29. Studnicka, G.M., Rahn, G.M., I.W. Cummings and W.A. Salser, 1978. Computer method predicting secondary structure of sing-stranded RNA. Nucleic Acids Res., 5: 3365-3387. http://nar.oxfordjournals.org/cgi/content/abstract/5/9/3365
30. Abrahams, J.P., M. Berg, E. Batenburg and C. Pleij, 1990. Prediction of RNA secondary structure, including pseudoknotting by computer simulation. Nucleic Acids Res., 18: 3035-3044. http://cnx.org/content/m11065/latest/
31. Van Batenburg, F.H.D., A.P. Gultyaev and C.W.A. Pleij, 1995. An APL-programmed genetic algorithm for the prediction of RNA secondary structure. J. Theor. Biol., 174: 269-280. DOI: 10.1006/jtbi.1995.0098
32. Gultyaev, A.P., F.H.D. Van Batenburg and C.W.A. Pleij, 1995. The computer simulation of RNA folding pathways using a genetic algorithm. J. Mol. Biol., 250: 37-51. DOI: 10.1006/jmbi.1995.0356
33. Shapiro, B.S. and J.C. Wu, 1997. Predicting RNA H-type pseudoknots with the massively parallel genetic algorithm. CABIOS., 13: 459-471. http://direct.bl.uk/bld/PlaceOrder.do?UIN=030659053&ETOC=RN&from=searchengine

34. Rivas, E. and S.R. Eddy, 1999. A dynamic programming algorithm for RNA structure prediction including pseudoknots. J. Mol. Biol., 285: 2053-2068. http://www.ncbi.nlm.nih.gov/pubmed/9925784
35. Dirks, R.M. and N.A. Pierce, 2003. A partition function algorithm for nucleic acid secondary structure including pseudoknots. J. Comput. Chem., 24: 1664-1677. http://www.ncbi.nlm.nih.gov/pubmed/12926009
36. Ruan, J.R., G.D. Stormo and W. Zhang, 2004. An iterated loop matching approach to the prediction of RNA secondary structures with pseudoknots. Bioinformatics, 20: 58-66. DOI: 10.1093/bioinformatics/btg373
37. Ren, J., B. Rastegari, A. Condon and H.H. Hoos, 2005. HotKnots: Heuristic prediction of RNA secondary structures including pseudoknots. RNA 11: 1494-1504. DOI: 10.1261/rna.7284905
38. Li, H., 2008. Heuristic Algorithm for Pseudoknotted RNA Structure Prediction. Proceedings of the 4th International Conference on Natural Computation, Oct. 18-20, IEEE Xplore Press, Jinan, pp: 542-546. DOI: DOI: 10.1109/ICNC.2008.676
39. Deogun, J.S., R. Donis, O. Komina and F. Ma, 2004. RNA secondary structure prediction with simple pseudoknots. Proceedings of 2nd Asia-Pacific Bioinformatics Conference, (APBC'04), Australian Computer Society, inc., Dunedin, New Zealand, pp: 1-8. http://crpit.com/confpapers/CRPITV29Deogun2.pdf
40. Reeder, J. and R. Giegerich, 2004. Design, implementation and evaluation of a practical pseudoknot folding algorithm based on thermodynamics. BMC. Bioinformat., 5: 104. DOI: 10.1186/1471-2105-5-104
41. Li, H. and D. Zhu, 2005. A New Pseudoknots Folding Algorithm for RNA Structure Prediction. Lecture Notes Comput. Sci., 3595: 94-103. http://cat.inist.fr/?aModele=afficheN&cpsidt=17096426
42. Huang, X. and H. Ali, 2006. High sensitivity RNA pseudoknot prediction. Nucleic Acid Res., 35: 656-663. DOI: 10.1093/nar/gkl943
43. Needleman, S.B. and C.D. Wunsch, 1970. A general method applicable to the search for similarities in the amino acid sequences of two proteins. J. Mol. Biol., 48: 443-453. http://www.ncbi.nlm.nih.gov/pubmed/5420325
44. Chowdhury, R.A., H. Le and V. Ramachandran, 2007. Efficient cache-oblivious string algorithms for Bioinformatics. Technical Report TR-07-03, Dept. of Computer Sciences, UT-Austin. ftp://ftp.cs.utexas.edu/pub/techreports/tr07-03.pdf
45. Chowdhury, R.A., H. Le and V. Ramachandran, 2008. Cache-oblivious dynamic programming for bioinformatics. Department of Computer Sciences, University of Texas at Austin, IEEE/ACM, http://www.cs.utexas.edu/~vlr/papers/tcbb08.pdf
46. Jabbari, H., A. Condon, A. Pop, C. Pop and Y. Zhao, 2007. HFold: RNA pseudoknotted secondary structure prediction using hierarchical folding. Lecture Notes Comput. Sci., 4645: 323-334. DOI: 10.1007/978-3-540-74126-8_30
47. Chowdhury, R.A. and V. Ramachandran, 2008. Cache-efficient dynamic programming algorithms for multicores. Proceedings of the 20th Annual Symposium on Parallelism in Algorithms and Architectures, June 14-16, ACM Press, Munich, Germany, pp: 207-216. http://doi.acm.org/10.1145/1378533.1378574